19. (a) Illuminate the architecture of Apache Hive with a neat diagram.

(b) Explain Serialization and Deserialization (SerDe) in Hive. Why is it important?

20. Elucidate the importance of Eval Functions and Complex Data Types in Pig.

APRIL/MAY 2025

23PCA42 — BIG DATA ANALYTICS

Time : Three hours Maximum : 75 marks

Shanmuga Industries T.V.Malai College

SECTION A — (10 × 2 = 20 marks)

Answer ALL questions.

1. Define structured data with an example.
2. List two challenges associated with Big Data.
3. What is NoSQL?
4. Mention the components of the Hadoop Ecosystem.
5. Name the basic data types in MongoDB.
6. State the role of a Mapper in MapReduce.
7. Name any four data types supported by Hive.
8. Write short notes on Serialization and Deserialization in Hive.

9. Delineate the key features of Pig.

10. What is the purpose of the DIAGNOSTIC operator in Pig?

SECTION B — (5 × 5 = 25 marks)

Answer ALL questions.

11. (a) Explain the classification of digital data with examples.

Or

(b) Illuminate the importance of Big Data Analytics in different industries.

12. (a) Compare Hadoop with traditional RDBMS.

Or

(b) Enlighten the process of reading and writing data in HDFS.

13. (a) Explicate the key differences between MongoDB and RDBMS.

Or

(b) What is the function of the Combiner and Partitioner in MapReduce? Explain with examples.

14. (a) What is partitioning in Hive? Explain its advantages and disadvantages with examples.

Or

(b) Explain Views in Hive. How are they different from tables?

15. (a) Give explanation the Anatomy and Features of Apache Pig.

Or

(b) Describe the execution modes of Pig and FOPS commands used in Pig.

Shanmuga Industries Arts & Science College T.V.Malai

SECTION C — (3 × 10 = 30 marks)

Answer any THREE questions.

16. Illustrate the evolution of Big Data and its importance in today's world.

17. Explain the architecture of Hadoop in detail.

18. Elaborate the MapReduce progamming model with an example.